

What is claimed is:

1. (canceled) rewritten/re-presented in claim 5
2. (canceled) rewritten/re-presented in claim 6
3. (canceled) rewritten/re-presented in claim 7
- 5 4. (canceled) rewritten/re-presented in claim 8
5. (canceled) rewritten/re-presented in claim 9
6. (canceled) rewritten/re-presented in claim 10
7. (canceled) rewritten/re-presented in claim 11
8. (canceled) rewritten/re-presented in claim 12
- 10 9. (re-presented - part of formerly independent claim 5) A method of evaluating whether an observed sequence of speech, image strip, or proteins has a subsequence being generated by one of a set of Hidden Markov Models (HMMs), comprising:
 - a) preprocessing the observation with any standard technique (like LPC or MFCC for utterances, choosing the section to be analyzed for images and proteins) to obtain
 - 15 a sequence X (which is a temporal sequence of speech feature vectors, respectively a linear spatial sequence of features for proteins and sections in images);
 - b) selecting a set of candidate patterns (like keywords, objects, respectively protein sequences for which we want to verify the existence in the current observation)

represented as hidden Markov models (HMMs) with a filler state q_G at beginning and one at the end;

c) selecting a confidence measures of the matching between a pattern and a subsequence of the observation, chosen among the following three ones:

- 5 c1) the accumulated posterior, normalized with the length of the matched subsequence X_b^e , b being the index of the first and e the index of the last element of the subsequence in X (aka. simple normalization)

$$\frac{-1}{e - b + 1} \log P(Q|X_b^e)$$

- 10 c2) a value obtained by partitioning the HMM states into subsets called phonemes, defined by a method $\text{Phonemes}(Q)$ that returns the segmentation of a path Q in the HMM into subsets of contiguous states, each subset belonging to a distinct phoneme, and computing one of:

- c2a) the worst average match in a phoneme, called real fitting,

$$\underset{Q}{\operatorname{argmin}} \left(\max_{Q \in \text{Phonemes}(Q)} \frac{\sum_{q^k \in Q} -\log P(q^k|x_k)}{|\{k|q^k \in Q\}|} \right)$$

- 15 c2b) double normalization of the accumulated posterior over the number of phonemes, J , and over the number of acoustic samples, $e_j - b_j + 1$, where e_j is the time frame where Q enters phoneme j , and b_j is the exit time frame from phoneme, j ,

$$\frac{-1}{J} \sum_{j=1}^J \left(\frac{1}{e_j - b_j + 1} \sum_{n=b_j}^{e_j} \log P(q_j^n|x_n) \right)$$

d) selecting a number called threshold, selection that can be done by user according to her experience with her application and environment;

e) computing for each candidate HMM pattern a number called 'score for the best matching with a subsequence of the observation',

5 by using a process that considers emission probabilities of x_G as zero, generating iteratively for each pair $\langle x_i, q^j \rangle$ between an element x_i of X and a state q^j of the current HMM, in the order of increasing i , a set of possible alternative paths in the HMM, that end in q^j and generate X_1^i ;

10 this set of paths being obtained by extending the paths associated with all the pairs $\langle x_{i-1}, q^k \rangle$ containing the previous element of X , (the empty path at x_1), and extended with transitions allowed by the analyzed HMM, each path being Q recorded by storing the length spent by Q in each phoneme (equivalent to storing the indexes of X where Q enters and exits the different phonemes), updating the previously chosen confidence measure of the obtained path,

15 and pruning the sets according to rules based on these confidence measures, namely where:

i) the simple normalization confidence measure is used with a safe pruning that discards a path Q_1 given the existence of an alternative path Q_2 in the same set, whenever $S_2 < S_1$ and $L_1 < L_2$, where S_1 and L_1 respectively S_2 and L_2 are the minus of the cumulated log of posteriors along the paths, and the lengths of the paths, for the paths Q_1 respectively Q_2 , and where the comparison is optionally optimized by sorting competing paths based on their

20

cost according to a merge-sort procedure;

ii) the double normalization confidence measure, on HMMs where no path skips any phoneme, is used with a safe pruning that discards a path Q_1 given the existence of a path Q_2 whenever one of the following tests succeeds:

- 5 (a) $l_2 \geq l_1$, $A > 0$, $B \leq 0$ and $L_c^2 A + L_c B + C \geq 0$
- (b) $l_2 \geq l_1$, $A \geq 0$, $B \geq 0$ and $C \geq 0$
- (c) $l_2 \geq l_1$, $A \leq 0$, $C \geq 0$ and $L^2 A + LB + C \geq 0$
- (d) $l_2 \geq l_1$, $A = 0$, $B < 0$ and $LB + C \geq 0$

10 where we denote by a_1 , p_1 , l_1 , respectively by a_2 , p_2 and l_2 the confidence measure for the previously visited phonemes, the posterior in the current phoneme and the length in the current phoneme for the path Q_1 , respectively the path Q_2 , and we also use the notations $A = a_1 - a_2$, $B = (a_1 - a_2)(l_1 + l_2) + p_1 - p_2$, $C = (a_1 - a_2)l_1 l_2 + p_1 l_2 - p_2 l_1$, $L = L_{max} - \max\{l_1, l_2\}$, $L_c = -B/2A$ and L_{max} is the maximum acceptable length for a phoneme;

15 and where each set of paths may optionally be reduced by storing only the best K matches;

iii) the double normalization confidence measure, on HMMs where some paths skip phonemes, is used with a safe pruning that discards a path Q_1 given the existence of a path Q_2 whenever $l_2 \geq l_1$, $A \geq 0$, $p_1 \geq p_2$ respectively $F_2 \geq F_1$,
 20 where F_1 respectively F_2 are the number of visited phonemes for paths Q_1 and Q_2 ;

iv) the real fitting is used with the safe pruning: Q_2 is discarded in favor of

another path Q_1 if the confidence measure of the Real Fitting for the previous phonemes is inferior (higher in value) for Q_2 compared with Q_1 , and if $p_1 \leq p_2$ and $l_2 \leq l_1$,

where p_1 , l_1 , respectively p_2 , l_2 represent the minus of the logarithm of the cumulated posterior respectively the number of frames in the current phoneme for the path Q_1 respectively Q_2 ,

and besides the previously mentioned safe pruning, heuristic prunings are also used for removing paths when $p > L_{max}P_{max}$ in any state or when $\frac{p}{l} > P_{max}$ at the output from a phoneme, where p and l are the values in the current phoneme for the minus of the logarithm of cumulated posterior and for the length of the path that is discarded;

and where each set of paths may optionally be reduced by storing only the best K matches;

f) returning as result the pattern with the highest score together with the score, or the set of all patterns with scores higher than the threshold, optionally with the boundaries of the subsequences of X that yield these scores.

10. (re-presented - part of formerly independent claim 5) A method of evaluating whether an observed sequence of speech, image strip, or proteins has a subsequence being generated by one element of a set of Hidden Markov Models (HMMs), comprising:

a) preprocessing the observation with any standard technique (like LPC or MFCC for utterances, choosing the section to be analyzed for images and proteins) to obtain

a sequence X (which is a temporal sequence of speech feature vectors, respectively a linear spatial sequence of features for proteins and sections in images);

b) selecting a set of candidate patterns (like keywords, objects, respectively protein sequences for which we want to verify the existence in the current observation) represented as hidden Markov models (HMMs);

c) selecting a number called threshold, selection that can be done by user according to her experience with her application and environment;

d) computing for each candidate HMM pattern a number called 'score for the best matching with a subsequence of the observation', or at least the information about whether this score is higher or lower than the threshold,

by using a method that applies Viterbi decoding for a HMM obtained by extending the initial one with a filler state just after start and one just before the termination state, and estimates the emission probability of the filler states in an iterative manner as being equal to

$$\frac{-1}{e - b + 1} \log P(Q^* | X_b^e)$$

for the path Q^* with highest score found in the previous iteration, where b and e being the indexes of X between which Q^* visits the HMM of the pattern,

and where the emission probability in the filler states in the first iteration can be initialized to any floating point number, but the iteration stops:

i) at convergence yielding the estimation of the boundaries and score of non-filler states of the HMM,

ii) when the confidence measure descends under a threshold value, T , estimating only the existence of a subsequence generated by the HMM,

iii) when the emission probability of filler states, ε_0 is initialized with T and is reestimated, as value of ε_1 at the end of the first iteration, to be higher than

5 T deciding that no subsequence was generated by the HMM,

e) returning as result the pattern with the highest score that is higher than the threshold, or the set of all patterns with scores higher than the threshold, optionally with the boundaries of the subsequences of X that yield these scores.

11. (re-presented - formerly dependent claim 8) The method of claim 9, where it carries out
10 the estimation of the existence of objects and their position in images, characterized by the fact that

the HMM patterns are built by describing sections through views of objects,
the emission probabilities are computed as a distance between colors (as a Gaussian with median at the color of the pattern, or a normalized inverse of the
15 Euclidean distance in the RGB space),

wherein the Hidden Markov Models that model the objects can be structured of distinct regions, that play in the frame of the method the role of the phonemes,
and wherein the properties of the transitions of the HMM models of the objects are optionally modified in a dynamic manner for each path during decoding (existence
20 and probability) by increasing/decreasing transition probabilities returning to the same state when matches in 'phonemes' were longer/shorter in the path than the

average predicted by the pattern.

12. (re-presented - formerly dependent claim 8) The method of claim 10, where it carries out the estimation of the existence of objects and their position in images, characterized by the fact that

5 the HMM patterns are built by describing sections through views of objects,

 the emission probabilities are computed as a distance between colors (as a Gaussian with median at the color of the pattern, or a normalized inverse of the

 Euclidean distance in the RGB space),

 and wherein the properties of the transitions of the HMM models of the objects are
10 optionally modified in a dynamic manner for each path during decoding (existence
 and probability) by increasing/decreasing transition probabilities returning to the
 same state when matches in 'phonemes' were longer/shorter in the path than the
 average predicted by the pattern.